Search for heavy metastable particles decaying to quark pairs in $p\overline{p}$ collisions at $\sqrt{s} = 1.96\ TeV$

The CDF Collaboration
URL http://www-cdf.fnal.gov
(Dated: November 29, 2010)

We report on an analysis searching for heavy particles that are metastable and decay into quark pairs with a macroscopic lifetime ($\sim 1\ cm$) at a displaced vertex. We use a data driven background approach, where we build probability density functions to model Standard Model secondary vertices from known processes in order to estimate the background contribution from the Standard Model. No statistically significant excess is observed above the background. Limits on the production cross section in a Hidden Valley benchmark phenomenology are set for various Higgs boson masses as well as metastable particle masses and lifetimes.

I. INTRODUCTION

We present an analysis describing the search for heavy metastable particles that decay at a displaced vertex away from the beamline. The Standard Model (SM) provides relatively light metastable particles such as B hadrons, Kaons, and Λs. However, there are no high mass particles with long lifetime. Our search is for a particle that originates from the primary interaction vertex, travels for a macroscopic distance ($\sim 1\ cm$) and then decays into quarks, which then hadronize into jets in the detector.

A recent analysis from D0 searched for heavy particles decaying with a displaced vertex, in this case into two b-quarks [1]. That analysis used the same phenomenological model as our analysis, the Hidden Valley model, which will be discussed in more detail in Section II. In addition, at CDF we employ the Silicon Vertex Trigger (SVT). The SVT allows CDF to trigger on tracks which originate from displaced decays. This ability allows us to increase signal efficiency while reducing the QCD background present at a hadron collider. It also makes out search sensitive to decays into any jets, not only b-jets, since we do not employ a muon trigger.

CDF is a general-purpose detector that is described in detail in Ref. [2]. The detector components relevant to this analysis are briefly described. Closest to the beampipe are multi-layer silicon detectors (SVX) [3]. These provide precision tracking which will be used to identify displaced vertices. Outside the SVX is an open-cell drift chamber covering the pseudorapidity region $|\eta| < 1$ [4]. (The pseudorapidity η is defined as $-\ln(\tan(\theta/2))$, where θ is the polar angle realtive to the proton beam direction.) The COT is used to recontruct charged particles' momenta. The tracking system is enclosed in a superconducting solenoid, which in turn is surrounded by a calorimeter. The CDF calorimeter system is separated into electromagnetic and hadronic subdetectors segmented in a projective tower geometry covering the region $|\eta| < 3.6$. The electromagnetic calorimeters utilize a lead-scintillator sampling technology [5], whereas the hadron calorimeters use iron-scintillator technology [6] [7]. Jets are reconstructed from the energy deposisted in these calorimeters [8]. Finally, the muon subdetectors are arrayed outside of the calorimeters. The beam luminosity is determined with gas Cherenkov counters located in the region $3.7 < |\eta| < 4.7$ which measure the average number of inelastic $p\overline{p}$ collisions per bunch crossing [10].

II. THE HIDDEN VALLEY MODEL

While the analysis presented here is a search for general particles that decay into jet pairs at a displaced vertex, a phenomenological theory is useful for a benchmark. The Hidden Valley (HV) phenomenology provides a framework in which we can generate signal MC, search for discriminants, optimize our search, and compare results. Here a brief outline of the Hidden Valley phenomenology is presented.

In the Hidden Valley phenomenology, the Standard Model(SM) gauge group G_{SM} is extended by a non-abelian group G_v [12] [11]. The canonical SM particles are neutral under G_v, but this new G_v contains light particles that are charged under G_v but neutral under G_{SM}. Heavy SM particles may carry both G_{SM} and G_v charge, which allow for interactions between SM fields and the new light particles. These would appear at the TeV scale at hadron colliders such as the Tevatron.

These v-particles would confine themselves inside v-hadrons, which in effect obey a v-QCD. Since the lightest v-hadrons are heavier than most SM particles, these v-hadrons can decay to SM particles. A wide range of masses,

lifetimes, and final states exist within the HV framework. If the lightest available HV particle, a v-π, equivalent to the SM π, has mass less than that of the top-quark, the predominant decay would be to $b\bar{b}$ quark pairs.

The lifetimes and masses of these v-hadrons are not constrained by the model [13]. With a non-zero lifetime, some particles would travel a distance from the primary vertex before decaying, much like a B or D hadron. This displaced vertex would be a signature to search for.

The HV also provides a way to search for the Higgs boson. If the Higgs mixes with this G_v field then it may decay to two (or more) v-hadrons. These v-hadrons would in turn decay into two $b\bar{b}$ quark pairs. It would be feasible to search for the Higgs boson using this final state at CDF. A Feynmann diagram of this decay is shown in Figure 1.

Although we have associated a phenomenological model, the HV model, this search can be for any general process containing heavy particles with a long-lifetime that decays with a displaced vertex away from the primary vertex. Results presented for this benchmark process can be used to constrain other models by accounting for the difference in the kinematic properties of the final state.

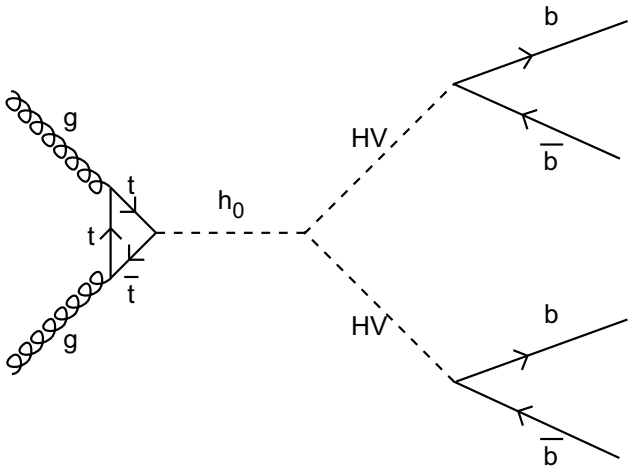

FIG. 1: Feynmann diagram of a basic Hidden Valley decay.

The Pythia MC program version 2.6 is used to generate the events for the signal MC [14]. In addition, CDF Tune A is included. GEANT is used for the detector simulation. The main process generated is $h_0 \to a^0, a^0 \to b\bar{b}, b\bar{b}$. SM Higgs production is used, as per the Hidden Valley model, $gg \to h_0$. However, the Higgs is constrained to decay into the HV particle, $h_0 \to a^0 a^0$. Here the a^0s are not the MSSM Higgs partners, but simply represent the HV particle. This allows us to generate the signal MC using the Pythia MC generator without significant modification. The a^0 has its mass artificially set to a HV particle mass, and its lifetime set to a HV particle lifetime. Finally, the a^0 has its decay daughters constrained to $b\bar{b}$ quark pairs.

Two Higgs masses are generated, one at relatively low mass, 130 GeV [15], and one at high mass, 170 GeV. Multiple HV particle masses from 20 GeV to 65 GeV are produced. The lifetime of the HV particle $c\tau_{HV}$ is set to 1.0 cm. Some signal MC samples have been weighted to study the lifetimes of the HV particle at 0.3, 2.5, or 5.0 cm.

III. EVENT SELECTION

A. ZBB Trigger

$Z \to b\bar{b}$ events are collected at CDF for the purposes of studying the Jet Energy Scale (JES) of b-quark jets [16]. A dedicated trigger has been designed to collect such events. The ZBB trigger path collects events where there are tracks with large impact parameter with respect to the primary vertex (d_0), i.e. displaced tracks. Because the Hidden Valley particle will decay at a displaced vertex, tracks from this decay have large d_0. Thus we use this trigger path in our analysis.

At Level-1, the trigger has two parts. It selects events with one central calorimeter tower with $E_T > 5\ GeV$ [17], and two tracks, one track with $p_T > 5.48\ GeV$, the other one with $p_T > 2.46\ GeV$.

At Level-2, there are two different subpaths that the trigger can take, named opposite side (OS) and same side (SS), which refer to the topological configuration of the displaced tracks in the event. The OS path consists of

Level-1	one central tower with $E_T > 5\ GeV$; two tracks: one track with $p_T > 5.48\ GeV$, one with $p_T > 2.46\ GeV$
Level-2	veto clusters with $E_T > 5\ GeV$, $1.1 < \|\eta\| < 3.6$; requires central/plug jet seed of 3 GeV; requires two clusters $E_T > 5\ GeV$, $\|\eta\| < 1.1$, which have $9 < \Delta Wedge < 12$; two SVT tracks with $p_T > 2\ GeV$, $160\mu m < \|d_0\| < 1000\mu m$, $\chi^2 < 12$
(OS)	$150° < \Delta\phi < 180°$
(SS)	$2° < \Delta\phi < 30°$
Level-3	two $\Delta R = 0.7$ jets with $E_T > 10\ GeV$, $\|\eta\| < 1.1$; two tracks with $p_T > 2\ GeV$, $\|\eta\| < 1.2$, $160\mu m < \|d_0\| < 1000\mu m$; two tracks with $p_T > 1.5\ GeV$ $\|\eta\| < 1.2$, $130\mu m < \|d_0\| < 1000\mu m$; IP significance $S(d_0) > 3$, $\Delta z < 5\ cm$

TABLE I: ZBB trigger path for trigger chunk 17. One of the two Level-2 paths (OS or SS) must be satified. There are two sets of track requirements at Level-3, the first is for SVT tracks, the second is for COT tracks.

three parts: a plug jet veto, a central calorimeter requirement, and displaced track requirement. The plug jet veto requires that there are no calorimeter clusters with $E_T > 5\ GeV$ in $|\eta| > 1.1$. The plug jet veto is designed to reduce the QCD background which produces more gluon radiation thatn the signal. The trigger requires two central calorimeter clusters, $E_T > 5\ GeV$ and $|\eta| < 1.1$, which are on opposite sides of the calorimeter. Finally, the track requirements at Level-2 are two SVT tracks which must point to one of the aforementioned clusters, with track $p_T > 2\ GeV$, $160\ \mu m < |d_0| < 1000\ \mu m$, $\chi^2 < 12$. The two tracks must have $150° < \Delta\phi < 180°$. The $|d_0|$ requirement selects displaced tracks. The SS path is similar to the OS path in that both have a plug jet veto and central calorimeter requirements. However, the SS path requires that the two displaced tracks point to a single cluster, and that $2° < \Delta\phi < 30°$.

At Level-3, the trigger requires two cone of 0.7 jets with $E_T > 10\ GeV$ and $|\eta| < 1.1$. It then also requires two tracks with $160\ \mu m < |d_0| < 1000\ \mu m$, simply verifying what was required at Level-2. As a cross-check this requirement is performed with both SVT tracks and COT tracks.

After the trigger selection, there is an additional event selection. First all the jets in this analysis must satisfy some basic requirements. Jets in this analysis are reconstructed with a $\Delta R = 0.4$ cone, where $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$. The E_T of the jet after Level-5 Jet Energy Scale (JES) [18] corrections must be greater than 20 GeV, $E_T^{L5} > 20$ GeV. The jets must be in the central region of the detector, $|\eta| < 1.0$. This requirement overlaps with the ZBB trigger requirement. These jets will be referred to as a tight-central jet.

B. Signal and Control Regions

While the Hidden Valley model predicts four jets in the final state, we select events with three jets in order to increase our acceptance. In addition, the plug-jet veto in the ZBB trigger at Level-2 reduces jet multiplicity (while simultaneously lowering the trigger rate at high luminosity). Thus in order to maintain acceptance the signal region is defined with three or more tight-central jets.

The signal MC samples generated show that the opening angle between the two jets is not back-to-back, but instead usually smaller. The effect of a lighter HV particle is more co-linear decay daughters. Thus for each dijet pair in a 3-or-more jet event, we apply a cut of $\Delta R < 2.5$ on each pair. Events which pass the jet multiplicity cut and have a dijet pair passing the ΔR cut are said be in the signal region.

In addition to the signal region, a two-jet control region is defined as follows. Events are required to have exactly two jets (again with $E_T^{L5} > 20\ GeV$) and a third jet with $E_T^{uncorr} < 15\ GeV$. There is no ΔR requirement on the two jets in this sample. The purpose of this control region is to allow us to test our background estimation technique on a set of events that is devoid of signal.

IV. BACKGROUND ESTIMATION

A. Secondary Vertex Tagging

Secondary vertex tagging (or b-tagging) is used in this analysis to identify jets with heavy flavor. SecVtx is the most commonly used b-tagging algorithm used at CDF. It is described in detail in Ref. [19]. Since heavy flavor quarks may decay at a displaced vertex, the SecVtx b-tagging algorithm is the starting point for the this displaced vertex search. However, we found that SecVtx had limitations which prompted us to clone this algorithm for our own use. TStnSVF is a program that is designed to perform b-tagging which performs the same general algorithm as SecVtx. It's functionality is that it allows the user to easily change the input parameters of the b-tagging algorithm.

Displaced tracks from the signal MC will be cut out by a maximum $|d_0|$ ($|d_0|_{max}$) cut placed on tracks by SecVtx, which requires tracks with a $|d_0|_{max} < 0.15\ cm$ with respect to the primary vertex. The SecVtx algorithm was designed mostly for top quark physics, where the daughter B hadron from the top decay originates at the primary vertex. The result of this cut is that the efficiency of b-tags in the signal MC is lower than expected. It is necessary to alter this $|d_0|_{max}$ cut to search for our signal, because B Hadrons that originate from a displaced vertex generate tracks with impact parameter larger than this value. By loosening this cut, we find that signal MC efficiency increases. However, this adjustment may result in a number of mistags.

Mistags are jets with a positive b-tag but which are not heavy flavor. Mistags arise from multiple sources, which include long lived Λs and K_Ss as well as nuclear interactions and other material effects within the detector material. Tracking errors and other mis-reconstruction also produce errant tracks that can be vertex as a mistag. Twenty $|d_0|_{max}$ cuts are chosen in this search to find a cut which maximizes the signal while minimizing the increase in mistags.

B. Modeling Background with Data

We want to model the SM backgrounds to the signal signature using real data. The ability to find particle's with displaced vertices relies on the reconstruction of secondary vertices. These secondary vertices can come from multiple SM sources.

Background	**SM Production**
b-quarks	QCD $b\bar{b}$, $t\bar{t}$, WZ/ZZ
c-quarks	QCD $c\bar{c}$, WZ/ZZ
light-flavor (mistags)	QCD $q\bar{q}$ & gg, hadronic τs

TABLE II: Standard Model processes that can produce particles that decay with a displaced vertex. This table is not-exclusive, but represents the majority of expected background processes.

The Monte Carlo simulation that is normally used to model backgrounds is inadequate when it comes to modeling these secondary vertices. One major concern is the effect that mistags have in this analysis. Mistags tend not to be modeled correctly in the Monte Carlo. One reason is the detector material is not fully modeled in the CDF detector simulation. By using a data-driven background we hope to accurately model what background exists in the signal data. In addition, this data driven background encapsulates different SM background sources, such as those listed in Table II. This eliminates the need to generate all SM background processes with MC.

1. Building p.d.f.s

The first step in modeling the background is to build standard model secondary vertex probability density functions (hereafter just p.d.f.s) of secondary vertices from b-tagged jets from background processes. These p.d.f.s involve multiple variables which characterize a secondary vertex within a jet. The p.d.f.s are constructed from data events, when possible, where the signal is not expected to be present, and in effect encapsulate standard model secondary vertex information.

The p.d.f. variables are defined in the plane transverse to the beamline, and all variables are two-dimensional. We define the secondary vertex's variables with respect to its parent jet's momentum vector, also called the jet axis. First, define $\vec{L_{xy}}$ as the two dimensional distance from the primary vertex to the secondary vertex. There are two components to this $\vec{L_{xy}}$ vector, one parallel to the jet axis, one perpendicular. These two components are the first two

p.d.f. variables, and are named u and v. These two variables define the position of a secondary vertex with respect to a jet axis.

The secondary vertex has a momentum associated with it. It is defined as the sum of the transverse momentum vectors of the tracks that are used to form the vertex. The $\Delta\phi$ angle between the jet axis and this momentum vector is the third p.d.f. variable, named α. This variable defines the direction of a secondary vertex with respect to a jet axis. We find that correlations exist between all three variables. Thus the resulting data structure for storing these p.d.f.s are three-dimensional histograms.

We split these p.d.f.s into three main categories for different quark flavors: b-quark, c-quark and light-flavor jets. These are further split into different bins of E_T^{L5} and are also split into three different categories of "number of SVT tracks." The ZBB trigger requires that there be two tracks that satisfy the trigger's SVT requirement within an event. This can be satisfied by two such tracks in one jet, or one track in each of two jets. Thus we split the jets into bins of 0, 1, and 2 or more SVT tracks, which would satisfy the ZBB trigger SVT requirements. As SVT tracks are displaced tracks, their presence is a likely indicator of heavy flavor decay. This binning, of E_T^{L5} and number of SVT tracks, is chosen because secondary vertex production changes with both E_T^{L5} and the number of SVT tracks, and thus the p.d.f.s' shapes are different in the different bins.

Different data sources are used to construct the different quark flavor p.d.f.s. We use a muon trigger with a relatively low p_T requirement to build the b-quark p.d.f.s. These data are rich in B Hadrons which decay semi-leptonically. To select events, we use a two jet selection where one jet is required to have a muon present within its jet cone, and both jets are positively b-tagged while being well separated in the detector. The non-muon jet, called the away-jet, is the jet of interest used to construct the b-quark p.d.f.s.

For the light-flavor p.d.f.s, we use the various QCD jet triggers which collect a large number of QCD multi-jet events. These data must have their heavy flavor contribution subtracted in order to isolate the light-flavor events with secondary vertices; in effect these are mistag p.d.f.s. The flavor composition of the jet triggers is calculated using QCD MC templates of b-quark, c-quark, and light-flavor secondary vertices. The tracks that form the secondary vertex have their four-momenta summed and the invariant mass calculated. Called the vertex mass, this quantity is distributed differently for secondary vertices from different flavor jets. We use a fraction fitter to fit these MC templates to the jet trigger data in order to determine the flavor composition. Afterwards, we subtract the b and c-quark contribution from the jet trigger p.d.f.s. The QCD MC templates are used as the source of the shapes of the b and c-quark p.d.f.s for this subtraction. Jets in the MC are matched to either a B hadron or D hadron in order to be identified as a b-quark or c-quark jet, respectively.

In order to ensure that the b and c-quark shapes for subtraction accurately represent the data, we compare the b-quark p.d.f.s from the muon trigger to the QCD MC p.d.f.s. where the jets are matched to a B hadron. Differences in the distributions of the p.d.f. variables are taken as a scale factor, which is first applied to both the b-quark and c-quark p.d.f. from the QCD MC before the subtraction is performed.

Finally, c-quark jets are not readily identifiable in real data. Therefore we turn to QCD MC in order to collect jets for the c-quark p.d.f.s.

2. Pseudo Event Generation

We use these p.d.f.s by applying them to multi-jet data, in which we are searching for the signal, to build a background estimate. While the kinematic information will come from the real event, the p.d.f.s are used to characterize secondary vertices of the real jets. These secondary vertices constructed from the p.d.f.s represent the background from the SM. When searching for our signal, we can use the kinematic information from the event, jet E_T^{L5}, η, etc. while constructing a secondary vertex that would represent processes from the SM, and not our signal. By comparing the events with these constructed secondary vertices, called "pseudo-events", with real data events we can search for the presence of new physics phenomenon.

The background estimate is created using the same data sample used to search for the signal. Events in the ZBB trigger are used for this construction. Only events in either the signal or control regions are selected, and separated such that there is a separate signal and control region background estimate. After events in the ZBB trigger are selected, they are sent through a pseudo event generation process.

With a dijet pair selected, from either the control region or signal region, we proceed to assign the jets in the pair pseudo b-tags. This is simply a flag, 0 or 1 indicating whether or not both jets have a (pseudo) positive b-tag. This pseudo b-tag is generated using the probability that dijets in this sample have (two) b-tags. We construct a dijet b-tag probability. This is the probability that **both** jets in the dijet have positive b-tags. The purpose of this dijet probability, as oppose to the probability of single jets being b-tagged, is to preserve kinematic correlations that may exist with respect to b-tagging.

Next the flavor of the pseudo jet is generated. The probability of a jet being a b-quark, c-quark, or light-flavor

jet is used to generate the flavor. However, unlike the flavor composition of the JET trigger samples where we were concerned with single jets, here we are concerned with the flavor composition of a pair of jets. With three possible flavor categories: b-quark, c-quark, and light-flavor, and two jets, there are nine possible combination of double-flavors for a pseudo-dijet: BB, BC, BL, CB, CC, CL, LB, LC, and LL; where mixed states such as the BC and CB states are not considered degenerate since the first flavor is for the more energetic of the two jets. This is done to preserve the correlations between the flavors of jets. Nature tends to create pairs of jets with the same underlaying flavor ($b\overline{b}$, $c\overline{c}$, etc.) which necessitates the need for two-dimensional fits.

We use two dimensional fits of the b-tag's vertex mass to determine the flavor of dijet pairs with positive b-tags. We use the same Pythia QCD dijet MC as previously described to build MC templates of the vertex mass. These templates are more complex now that we are dealing with a simultaneous fit. First the individual b-quark, c-quark, and light-flavor MC templates are joined to form two-dimensional vertex mass p.d.f.s for BB, BC, BL, etc. Then the vertex mass p.d.f.s are merged to form a single set of vertex mass p.d.f.s that encompasses all nine double-flavor states. Because the nine double-flavors are constrained by the physical property that they must add to one, there are eight fractions which we fit. These eight fractions are algebraic combinations of the nine double-flavor states. Fits are performed using the `RooFit` package [20].

Finally, the secondary vertex information is sampled from the p.d.f.s, generated using background processes. In this step the jets are sampled independently. The sampling is performed on the three-dimensional histogram where the p.d.f.s information is stored. A random (but correlated) u, v, and α are chosen and assigned to the pseudo-jet.

At the end of this procedure what we have created is a set of pseudo-events (made up of pseudo-dijets) where the b-tagging proportion is the same as the ZBB trigger data, the flavor of the pseudo-dijets is derived from the same ZBB trigger data, and the secondary vertex information is obtained from the SM via the p.d.f.s. A pseudo event contains all the kinematics of the real event it is generated from, but the secondary vertex information comes from the background processes via the p.d.f.s. We can then search for our signal using the real ZBB data, while using the pseudo-events to estimate the SM background from SM processes.

3. Validation

We use the control region to validate this algorithm. Because we expect the control region to be devoid of signal, we can compare the real dijet data to the pseudo-dijets generated to see if the pseudo-events are well behaved. For the purposes of this validation, exactly one pseudo-event was generated for each real event, and the p.d.f.s were only sampled once for each pseudo-jet. Distributions of the control region pseudo-event vs. real events show that the pseudo-event generation is well behaved. Figure 2 shows the distribution of ζ in the control region for real events and pseudo-events.

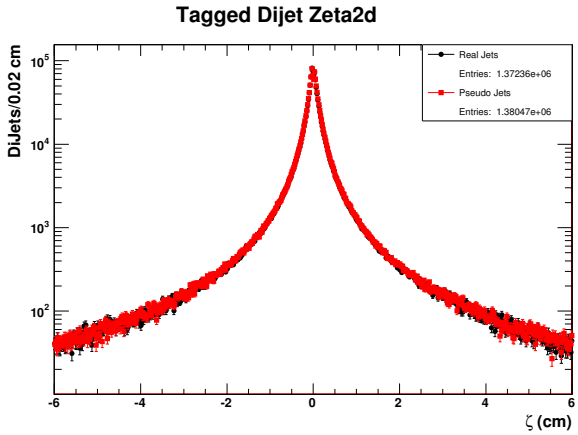

FIG. 2: ζ distribution for the real events (black) and pseudo-events (red) in the control region.

V. SIGNAL SEARCH

A. Discriminants from Signal MC

In our search for discriminants we look at multiple variables that would give us power to isolate the signal. However, none of the standard variables used in CDF analyses seem to suffice. Thus, we develop our own variables for this analysis, which take advantage of the displaced vertices present in our events.

Figures 3 and 4 show diagrams for the two new variables that we develop for this analysis: ψ and ζ. In both cases, the figures are in a two-dimensional plane transverse to the beamline. Jets with positive secondary vertices have a secondary vertex (a position) and the sum of the four momenta of the tracks that make up that vertex (a direction). With a position and a direction we can draw a line in two-dimensional space. We define $\vec{\psi}$, as the orthogonal vector from the primary vertex to this line. The magnitude of this vector, $\psi = |\vec{\psi}|$, is the distance of closest approach of this line to the primary vertex.

ψ can be a signed variable. The dot product of the jet's momentum vector and the ψ vector is calculated, and if positive ψ is positive; likewise ψ is negative if the dot product is negative. Although this variable can be signed, the ψ distribution is symmetric about zero, and the sign does not add to its discriminating power.

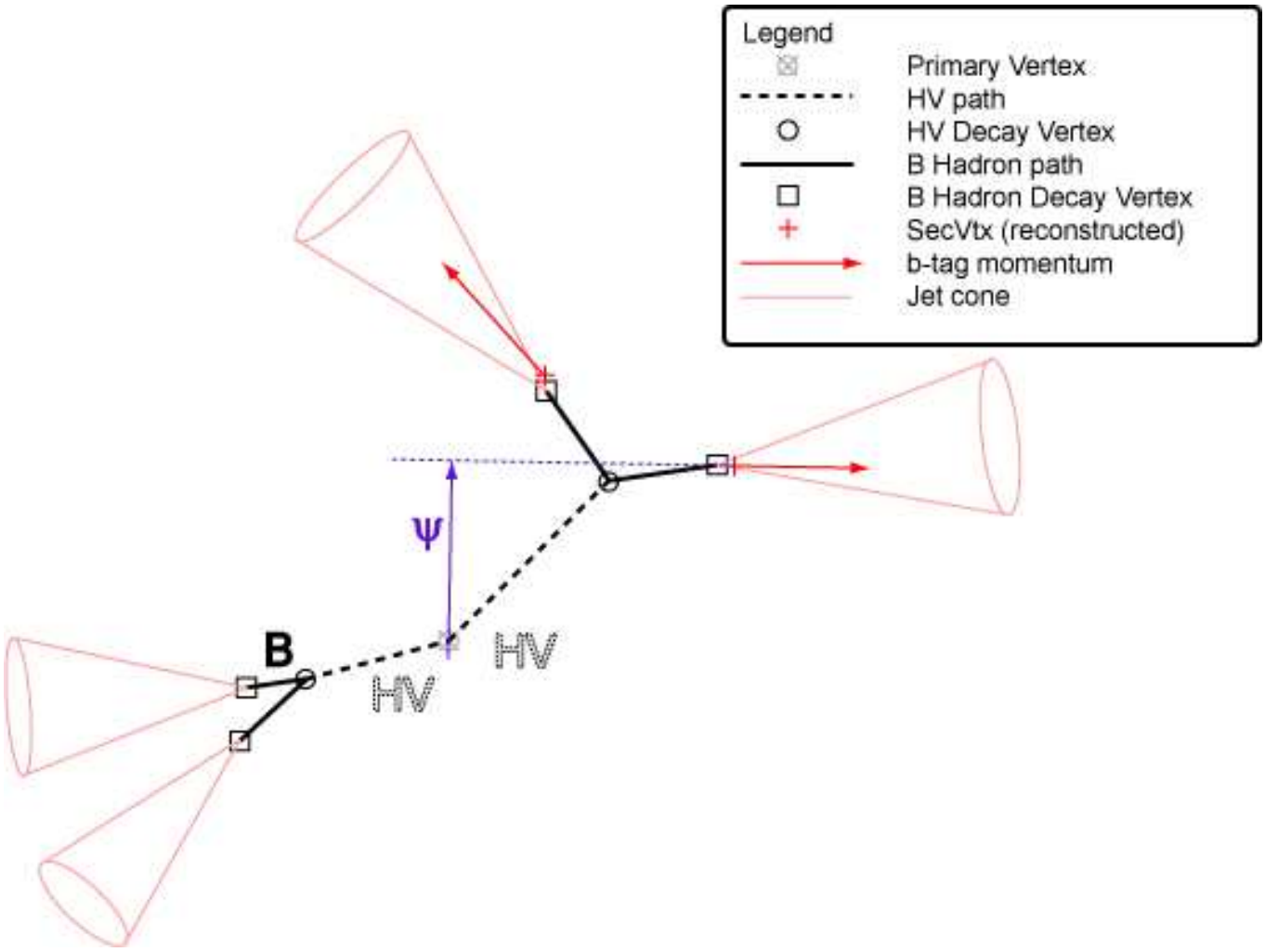

FIG. 3: Schematic diagram of variable ψ, the impact parameter of a jet with a positive b-tag/secondary vertex. This figure is not to scale. The figure is shown in a plane transverse to the beamline.

ζ is defined for events where there are two positive b-tagged jets. With two such jets, we have two vertex positions, and their corresponding directions, see Figure 4. We can construct two lines in two dimensional space. These lines must either be parallel, coincident, or intersect, and in nearly every case they intersect. This intersection can be thought

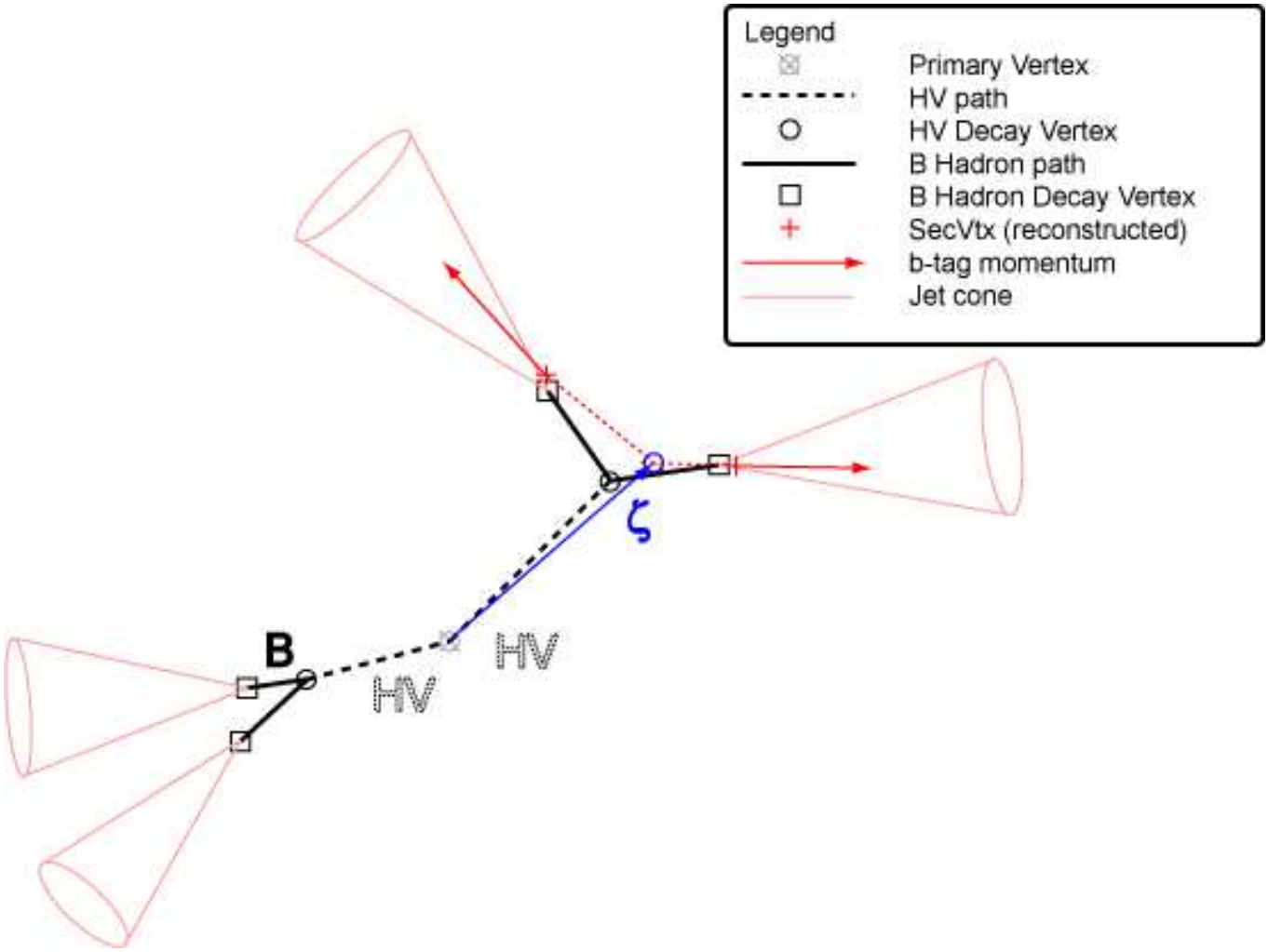

FIG. 4: Schematic diagram of variable $\vec{\zeta}$ which represents the decay vertex of the HV particle.

of as the reconstructed decay vertex of the HV particle. The vector from the primary vertex to this reconstructed decay vertex is $\vec{\zeta}$. And the magnitude $\zeta = |\vec{\zeta}|$ is the lifetime or decay distance of the HV particle.

ζ can be signed as well. The sum of the four momenta of the two jets with b-tags is calculated. The dot product of $\vec{\zeta}$ and this sum is then calculated. If the dot product is positive, then ζ is positive and likewise for a negative dot product. The sign effectively indicates whether or not the decay vertex is in the same hemisphere of the detector as the direction of the two jets. Signal MC events have more positive zeta than negative, while the background MC events have zeta uniformly distributed around zero.

B. From Pseudo-events to a Background Estimate

To generate the background estimate, we sample from the p.d.f.s many times for a given pseudo-dijet pair to build multiple pseudo-data. The first two of three steps of the pseudo-event generation described in Section IV are performed once for each real ZBB trigger event. But then the p.d.f. sampling step is performed many times, with each sampling considered a separate "pseudo-experiment". Each pseudo-experiment is then treated independently, and is passed through the same set of analysis cuts, which will be described in further detail below. The resulting number of events that pass these cuts is collected for each pseudo-experiment. The mean of this distribution is termed the background estimate.

The background estimate represents the number of events in data that would pass the analysis cuts if only SM processes contributed to the observed data. In effect it is the null hypothesis for this search. We perform a simple

counting experiment by comparing this null hypothesis to the number of observed data events with the same analysis cuts.

C. Signal Background Optimization

To conduct our search, we investigate the following variables, optimizing the last three.

1. $|d_0|_{max}$ cut of tracks that are used by the b-tagging algorithm
2. Angle between the two jets (ΔR)
3. $|\psi|$, the impact parameter of a b-tagged jet
4. ζ, the decay distance of the HV particle

The $|d_0|_{max}$ cut is a parameter of the b-tagging algorithm. Figure 5 shows the behavior of the signal MC and background pseudo-data, where one pseudo-event was generated for each real event, respectively, for the twenty $|d_0|_{max}$ cuts investigated.

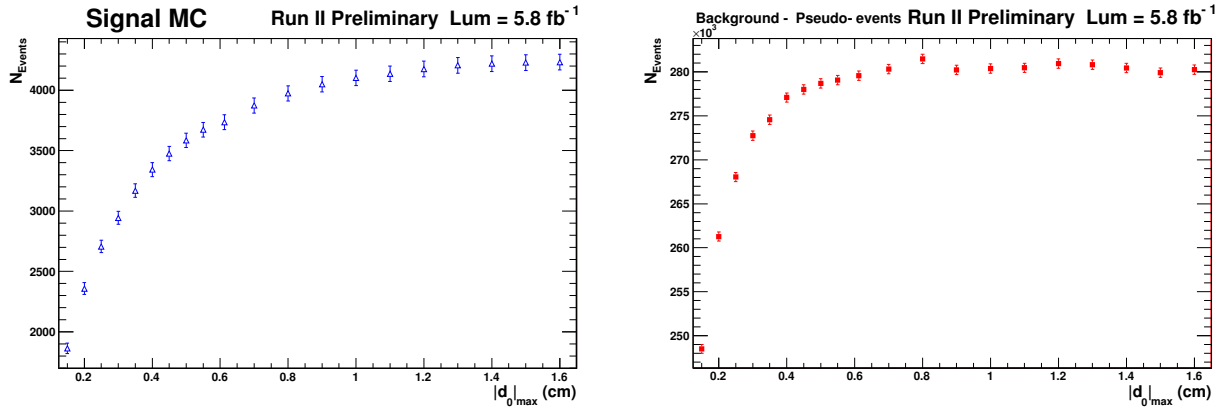

FIG. 5: Number of signal (left) and background (right) events vs. $|d_0|_{max}$.

At some point both distributions plateau. The number of events in the background sample jumps up and down. This is due to the effect that sometimes adding an additional track reduces the number of b-tags at that particular operating point. This occurs when the additional track contributes too much to the χ^2 of the vertex fit and cannot be pruned from the vertex, thus eliminating the b-tag altogether.

The distribution for the signal shows that larger $|d_0|_{max}$ cuts allow for more signal acceptance, while the background plateaus at about $|d_0|_{max} < 0.70\ cm$. Logically, this means we can pick a maximal $|d_0|_{max}$ cut, which would accept as many signal MC events as possible. However, this is not feasible as there are physical constraints present due to the CDF detector.

The CDF inner detector contains a beampipe with $r = 1.26\ cm$. Attached to the beampipe is a single layer of silicon strips, L00. While it is not required that tracks deposit hits in this inner-most layer, we want to ensure that tracks originating from the primary vertex *could* hit this detector. Thus the d_0 of a tracks originating from a primary vertex must have less than the radius of the beam pipe. However, the primary vertex at CDF is not at the exact center of the detector. Accounting for this shift, we determine that tracks with a maximum $|d_0| < 1.0\ cm$ may deposit hits in L00, thus a $|d_0|_{max} < 1.0\ cm$ is chosen.

With the $|d_0|_{max}$ cut set, we optimize the other three variables. The cuts chosen for the low and high HV mass searches are shown in Table III. The searches were optimized separately because the low mass HV results in daughter jets which are more co-linear. This changes the nature of the ΔR cut. For the low HV mass search, only a $\Delta R_{max} < 0.75$ cut is imposed; no ΔR_{min} cut is applied.

An additional cut is imposed on ζ. The length of ζ must be less than the distance from the primary vertex to the closest secondary vertex. This ensures that the intersection is "behind" both secondary vertices.

An unanticipated source of background became apparent when we applied these analysis cuts to the real ZBB trigger events. A few events in the low HV mass sample appear to contain a single secondary vertex from a B hadron,

Variable	high HV mass		low HV mass	
$\|d_0\|_{max}$ (cm)	$<$	1.0	$<$	1.0
ΔR_{min}	$>$	0.75		n.a.
ΔR_{max}	$<$	2.0	$<$	0.75
$\|\psi\|$ (both jets) (cm)	$>$	0.11	$>$	0.12
ζ (cm)	$>$	0.8	$>$	0.7

TABLE III: Variable cuts for both the low and high HV mass searches.

in which some of the decay products are found in each of two nearby jet cones. Two features of these events are that 1) the two secondary vertices are very close to each other, a variable named ΔS_{2d}, and 2) the total invariant mass of all the tracks in both vertices ($TotalVtxMass$) is less than the b-quark mass. These cuts were added to the low HV mass search analysis cuts because when these required were applied to the signal MC sample, there is a negligible reduction in the efficiency, see Table IV.

Variable	low HV mass	
ΔS_{2d} (cm) OR	$>$	0.06
$TotalVtxMass$ (GeV)	$>$	5.0

TABLE IV: Additional cuts on the low HV mass search due to an unanticipated background.

D. Results

With the variable cuts set, we proceed to run 10,000 pseudo-experiments for both mass searches. The results are shown in Figure 6. The low HV mass search (left) and high HV mass search (right) are Poisson distributions with means $\mu_{low} = 0.58$ and $\mu_{high} = 0.29$. These are the estimated numbers of SM background events. The statistical uncertainty on these numbers, the mean divided by the squared root of the number of pseudo-experiments, is negligible.

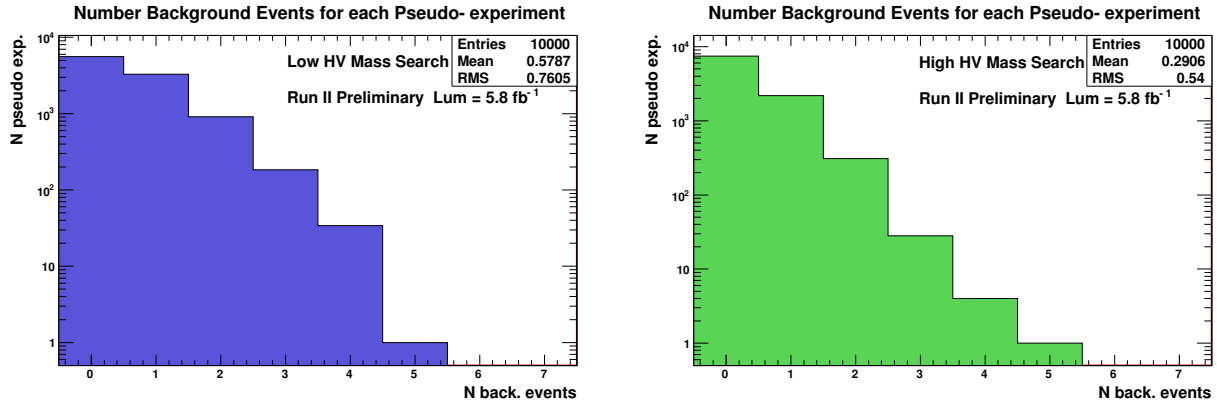

FIG. 6: Background event distributions for low (left) and high (right) HV mass searches. The mean of these distributions are the background estimates.

With the same variable cuts we can also calculate the number of expected signal MC events that we would expect with the same integrated luminosity as the ZBB trigger. This is done by calculating the number of events that pass the cuts in each signal MC sample and multiplying this number by a scale factor consisting of the luminosity of the ZBB trigger sample multiplied by the cross section for Higgs production ($gg \to h_0$) divided by the number of signal MC events generated. The Higgs cross sections were obtained from Ref. [22]. For $M_{h0} = 130\ GeV$ the cross section is $\sigma_{gg\to h0} = 858\ pb$, while for $M_{h0} = 170\ GeV$ the cross section is $\sigma_{gg\to h0} = 349\ pb$. The branching ratio of the Higgs is assumed to be 100% to the HV particles, and the branching ratio of the HV particles is assumed to be 100% to $b\bar{b}$

pairs. A B-Tagging Scale Factor (0.9) for the TStnSVF algorithm with $|d_0|_{max} < 1.0\ cm$ has also been applied twice, because we have two b-tagged jets.

Due to the unanticipated background mentioned earlier, additional cuts are applied to the number of expected signal MC calculation in the low HV mass search.

Table V shows the results of our search. The number of expected signal events and its statistical error is calculated from the signal MC samples. The number of background events is also shown, its uncertainty is calculated in the next section. There s one events observed in the low HV mass search and one in the high HV mass search.

Higgs Mass (GeV)	**HV Mass (GeV)**	**HV lifetime (cm)**	**Expected Signal MC**	σ_{stat}	**Background Estimate**	**Number Observed**
low HV mass search						
130	20	1.0	0.725	0.034	0.58	1
170	20	1.0	0.0856	0.0082	0.58	1
high HV mass search						
130	40	1.0	0.379	0.027	0.29	1
170	40	1.0	0.492	0.019	0.29	1
170	65	1.0	0.211	0.012	0.29	1
130	40	0.3	0.384	0.027	0.29	1
130	40	2.5	0.139	0.016	0.29	1
130	40	5.0	0.0489	0.0096	0.29	1

TABLE V: Results of our search. The background estimate uncertainties are discussed below.

VI. SYSTEMATIC UNCERTAINTIES

The systematic uncertainties in this analysis fall into two main categories. The first are systematics that affect the background estimate. The second are the systematics that affect the number of expected signal MC events.

In the first category, there are three major sources of uncertainty, each corresponding to a step in the pseudo-event generation. Statistical uncertainties arise from the b-tagging probability calculated to give events pseudo b-tags. These are propagated through as systematic uncertainties. The flavor composition probabilities used to determine the pseudo-flavor of the jets have two systematic uncertainties: the statistical uncertainty from the flavor composition fractional fit and a systematic due to the MC over-efficieny is track reconstruction, which has a direct impact on the vertex mass of the secondary vertex. We use an overall 3% reduction in the vertex mass to model a maximal variation this over-efficiency would produce.

For each systematic, we generate a new set of pseudo-data that represents the systematic shift. Then we perform another 10,000 pseudo-experiments as before and compare the background estimate to the central value calculated in Table V. The percent difference is used as the systematic.

For the last of the three steps in pseudo-event generation, the p.d.f. sampling, we turn to the bootstrap technique [21]. Effectively the three dimensional histograms are statistically varied within their Poisson statistical fluctuations, and a new background estimate is calculated using 10,000 pseudo-experiments. This procedure is itself replicated 200 times. The standard deviation of these 200 means is the uncertainty associated with the p.d.f. sampling. Figure 7 shows the results of this technique for both the low and high HV mass searches.

The second category of systematic uncertainties affects the signal estimate. These include:

1. Jet Energy Scale
2. Trigger Efficiency
3. B-tagging Scale Factor
4. Parton distribution function
5. Luminosity

The first two are systematic uncertainties that are calculated separately for each signal MC sample, while the BTSF and Luminosity have the same value across all samples, and the Parton distribution function is approximately the same value for all samples.

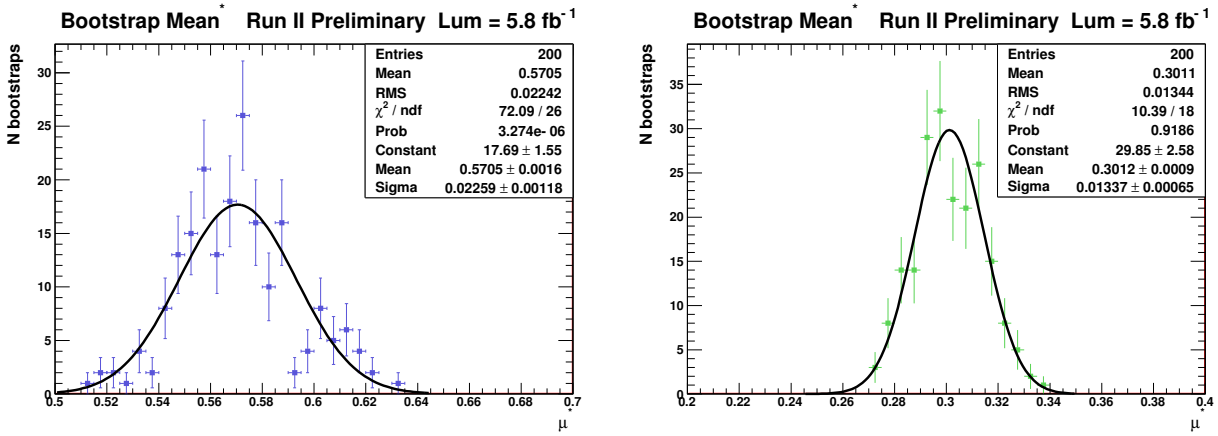

FIG. 7: Results of the bootstrap. The distributions show the results of 200 bootstrap where each entry consists of running 10,000 pseudo-experiments with a bootstrap-p.d.f. sample. The widths of the distributions are the uncertainties on the background estimate.

The JES factor on jets is varied up (down) one sigma with respect to its central value. The result is that more (less) jets pass the $E_T^{L5} < 20\ GeV$ cut. This effects the number of expected signal MC events for each sample. The trigger efficiency is a result of applying the ZBB trigger simulation on the generated events. Since we use triggered events to calculate the signal MC, the uncertainty in the trigger efficiency propagates as a systematic.

The BTSF systematic uncertainty for the TStnSVF b-tagger is calculated at the operating point $|d_0|_{max} < 1.0\ cm$. The parton distribution function uncertainty is taken from Ref. [22] which documents this uncertainty for multiple analyses, including ones that use $gg \to h_0$ production. Finally, the Luminosity uncertainty is taken from Ref. [23].

All the systematic uncertainties calculated are shown in Table VI. These systematic uncertainties are used in the calculation of the limits discussed in Section VII.

Uncertainty	**Down (%)**	**Up (%)**
Background estimite - low HV mass search		
PDF statistics	±0.039	
B-tag prob.	-7.74	3.42
Flavor Comp.	-0.5	2.75
Background estimate - high HV mass search		
PDF statistics	±0.046	
B-tag prob.	±3.92	
Flavor Comp.	-0.5	8.91
Signal MC		
JES	-15.6% to -6.3%	4.0% to 25.5%
Trigger Eff.	-1.1% to -0.41%	0.45% to 1.36%
BTSF	±10	
Parton dist. func.	±2.5	
Luminosity	±6	

TABLE VI: Summary of systematic uncertainties for the background estimate and signal MC. The JES and Trigger efficiency are calculated separatly for each signal MC sample.

VII. CONCLUSION

With all the uncertainties calculated we form test hypotheses consisting of our background estimate along with our signal MC. A separate test hypothesis is constructed for each signal MC sample. We also create corresponding null hypotheses consisting only of the background estimate for each HV mass search. No statistically significant signal is seen in our search. Table VII shows p-values for each signal MC sample and search, showing the probability that the

null hypothesis has fluctuated to the data.

Higgs Mass (GeV)	**HV Mass (GeV)**	**HV life-time (cm)**	**p-value**
low HV mass search			
130	20	1.0	0.48
170	20	1.0	0.49
high HV mass search			
130	40	1.0	0.29
170	40	1.0	0.28
170	65	1.0	0.29
130	40	0.3	0.29
130	40	2.5	0.30
130	40	5.0	0.29

TABLE VII: Null hypothesis p-values for this search.

Since we do not observe a statistically significant excess we proceed to set a limit on the production cross section times branching ratio of the Hidden Valley model for the particular masses and lifetimes we studied. A Bayesian limit calculator [24] is used for this calculation. Table VIII shows the resulting observed limit, median expected limit, along with the ± 1 and ± 2 sigma values on the expected limit, all at 95% confidence level.

Higgs Mass (GeV)	**HV Mass (GeV)**	**HV life-time (cm)**	**Obs. Limit (pb)**	**Expected Limit (pb)**				
				-2 σ	-1 σ	median	+1 σ	+2 σ
low HV mass search								
130	20	1.0	4.51	3.14	3.14	3.14	4.51	7.62
170	20	1.0	14.96	10.38	10.38	10.38	14.96	25.36
high HV mass search								
130	40	1.0	9.08	5.98	5.98	5.98	9.01	12.15
170	40	1.0	2.95	1.96	1.96	1.96	2.95	3.96
170	65	1.0	6.90	4.60	4.60	4.60	6.90	9.28
130	40	0.3	8.79	5.83	5.83	5.83	8.79	11.86
130	40	2.5	23.11	15.29	15.29	15.29	23.11	31.26
130	40	5.0	62.64	41.28	41.28	41.28	62.46	85.00

TABLE VIII: Limit calculation results for different signal MC samples.

The counting experiment was performed with a small discrete number of events, where the background estimate is less than one. Thus the expected limit can only fluctuate up (from 0). The result is that the negative sigma expected limits will be identical to the median limit.

Figures 8, 9, and 10 show the results of the limit calculation. In Figures 8 and 9 the x-axis is the mass of the Higgs boson. Figure 8 is for a M_{HV} of 20 GeV corresponding to the low HV mass search. Figure 9 is for a M_{HV} of 40 and 65 GeV, corresponding to the high HV mass search. Figure 10 shows the limits for M_{h_0} of 130 GeV and M_{HV} of 40 GeV with the HV lifetime on the x-axis.

We have searched for heavy metastable particles that decay into a jet pair at a displaced vertex at CDF. No statistically significant excess is observed, and limits are set on the production cross section times branching ratio for the Hidden Valley phenomenology we have used as a benchmark. The results shown for this phenomenology can be used to constrain other models considering the differences of the cross section, branching ratio, and the kinematics of the final state.

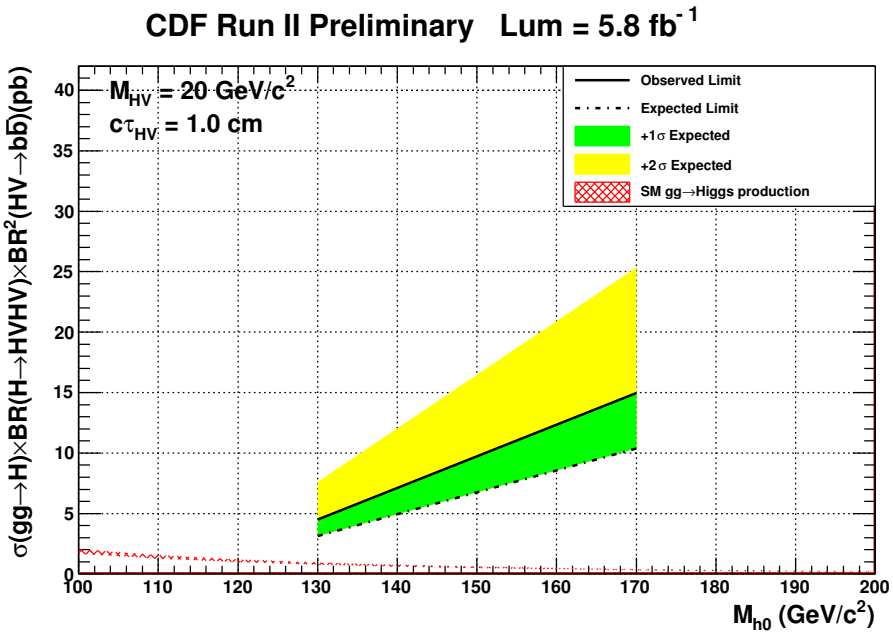

FIG. 8: Observed and Expected limit with +1 & +2 σ bands for signal MC with HV mass 20 GeV.

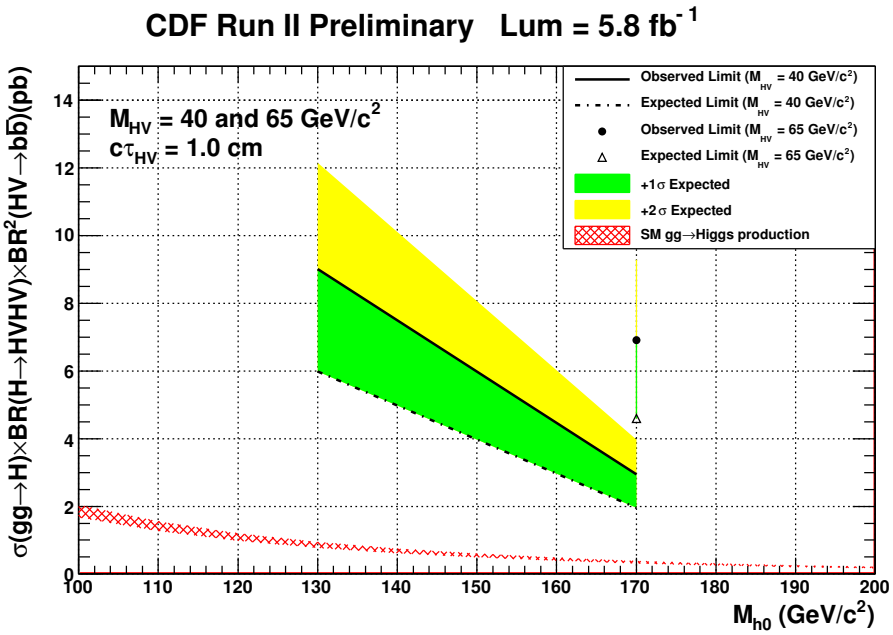

FIG. 9: Observed and Expected limit with +1 & +2 σ bands for signal MC with HV masses 40 and 65 GeV.

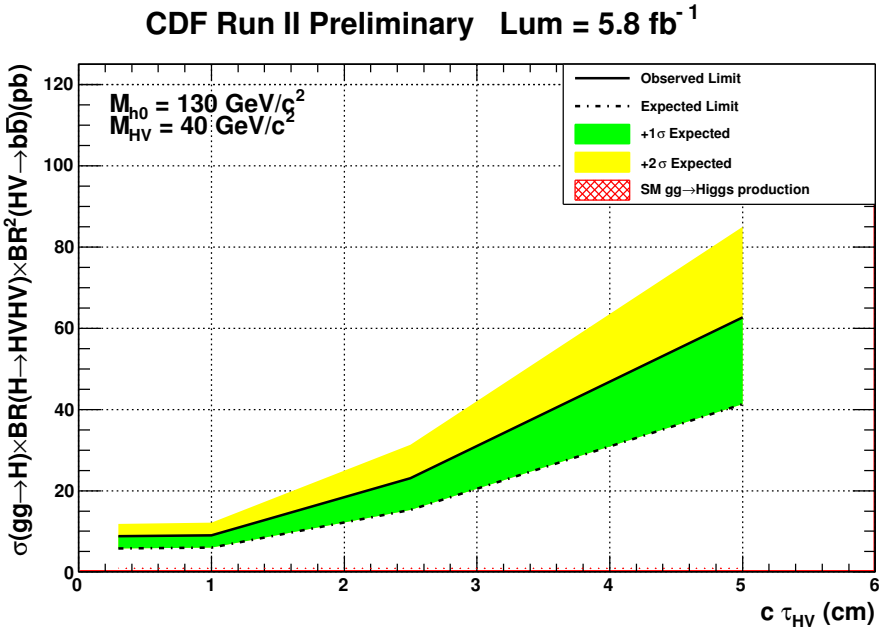

FIG. 10: Observed and Expected limit with +1 & 2 σ bands for signal MC for differing HV particle lifetimes.

Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium für Bildung und Forschung, Germany; the World Class University Program, the National Research Foundation of Korea; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland.

---

[1] V. M. Abazovm *et al.* Int. J. Mod. Phys., A **20**, 3263 (2005).
[2] F. Abe, *t al.*, Nucl. Instrum. Methods Phys. Res. A **271**, 387 (1988); D. E. Acosta, *et al.* (CDF Collaboration), Phys. Rev. D **71**, 052003 (2005); The CDF-II Detector Technical Design Report, Fermilab-Pub-96/390-E.
[3] A. Sill, *et al.*, Nucl. Instrum. Methods A **447**, 1 (2000); A. Affolder, *et al.*, Nucl. Instrum. Methods A **453**, 84 (2000); C. S. Hill, *et al.*, Nucl. Instrum. Methods A **530**, 1 (2000).
[4] A. Affolder, *et al.*, Nucl. Instrum. Methods A **526**, 249 (2004).
[5] L. Balka, *et al.*, Nucl. Instrum. Methods A **267**, 272 (1988); S. R. Hahn, *et al.*, Nucl. Instrum. Methods A **267**, 351 (1988).
[6] S. Bertolucci, *et al.*, Nucl. Instrum. Methods, A **267**, 301 (1988); M. Albrow *et al.*, Nucl. Instrum. Methods, A **480**, 524 (2002); G. Apollinari *et al.*, Nucl. Instrum. Methods, A **412**, 515 (1998).
[7] S. Kuhlmann, *et al.*, Nucl. Instrum. Methods A **518**, 39, 2004.
[8] F. Abe *et al.* (CDF Collaboration), Phys. Rev. D **45**, 1448 (1992).
[9] G. Ascoli *et al.*, Nucl. Instrum. Methods A **268**, 33 (1988).
[10] D. Acosta *et al.*, Nucl. Instrum. Methods A **494**, 57 (2002).
[11] M. Strassler and K. Zurek, Physics Letters B **661**, 263 (2008). hep-ph/0605193v2.
[12] M. Strassler and K. Zurek, Physics Letters B **651**, 374 (2007). hep-ph/0604261v2/
[13] M. Strassler (unpublished). See *http://www.phys.washington.edu/users/strasslr/hv/bench*
[14] T. Sjostrand, S. Mrenna, P. Skands, JHEP **05**, 026 (2006).
[15] We use the convention that "momentum" refers to GeV/c and "mass" to GeV/c^2.
[16] J. Donini *et al.*, Nucl. Instrum. Methods A **596**, 354 (2008). hep-ex/0801.3906.
[17] The CDF II uses a cylindrical coordinate system in which ϕ is the azimuthal angle, r is the radius from the nominal beamline, and z points in the proton beam direction. The transverse plane, $r-\phi$ plane, or sometimes referred to as x-y plane is the plane perpendicular to the z axis. Transverse momentum and energy are the projections of total momentum and energy onto the $r-\phi$ plane and defined as $\mathrm{p_T} = p\sin\theta$ and $\mathrm{E_T} = E\sin\theta$, respectively. Here, θ is the polar angle measured with respect to the interaction vertex.
[18] A. Bhatti *et al.*, Nucl. Instrum. Methods A **566**, 375 (2006). hep-ex/0510047.
[19] D. E. Acosta, *et al.* (CDF Collaboration), Phys. Rev. D **71**, 052003 (2005); C. Neu, FERMILAB-CONF-06-162-E.
[20] W. Verkerke and D. Kirkby (2003). For documentation and source code, see *http://roofit.sourceforge.net* and arXiv:physics/0306116.
[21] B. Efron and R. J. Tibshirani, "An Introduction to the Bootstrap". (New York : Chapman & Hall, 1993).
[22] The TEVNPH Working Group (CDF and D0 Collaborations) (2010), arXiv:1007.4587.
[23] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. Lett. 94, 091803 (2005).
[24] T. Junk (2007) For documentation and source code, see *http://http://www-cdf.fnal.gov/ trj/mclimit/production/mcliit.html* and CDF/DOC/STATISTICS/DOC/8128.